Data Assimilation in Reduced Modeling

Peter Binev, Albert Cohen, Wolfgang Dahmen,
Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk *

June 17, 2015

Abstract

This paper considers the problem of optimal recovery of an element u of a Hilbert space $\mathcal{H}$ from measurements of the form $\ell_j(u)$, $j = 1, \dots, m$, where the ℓ_j are known linear functionals on $\mathcal{H}$. Problems of this type are well studied [18] and usually are carried out under an assumption that u belongs to a prescribed model class, typically a known compact subset of $\mathcal{H}$. Motivated by reduced modeling for solving parametric partial differential equations, this paper considers another setting where the additional information about u is in the form of how well u can be approximated by a certain known subspace V_n of $\mathcal{H}$ of dimension n, or more generally, in the form of how well u can be approximated by each of a sequence of nested subspaces $V_0 \subset V_1 \cdots \subset V_n$ with each V_k of dimension k. A recovery algorithm for the one-space formulation was proposed in [16]. Their algorithm is proven, in the present paper, to be optimal. It is also shown how the recovery problem for the one-space problem, has a simple formulation, if certain favorable bases are chosen to represent V_n and the measurements. The major contribution of the present paper is to analyze the multi-space case. It is shown that, in this multi-space case, the set of all u that satisfy the given information can be described as the intersection of a family of known ellipsoids in $\mathcal{H}$. It follows that a near optimal recovery algorithm in the multi-space problem is provided by identifying any point in this intersection. It is easy to see that the accuracy of recovery of u in the multi-space setting can be much better than in the one-space problems. Two iterative algorithms based on alternating projections are proposed for recovery in the multi-space problem and one of them is analyzed in detail. This analysis includes an a posteriori estimate for the performance of the iterates. These a posteriori estimates can serve both as a stopping criteria in the algorithm and also as a method to derive convergence rates. Since the limit of the algorithm is a point in the intersection of the aforementioned ellipsoids, it provides a near optimal recovery for u.

Keywords: optimal recovery, reduced modeling, greedy algorithms
MSC numbers: 62M45, 65D05, 68Q32, 97N50.

*This research was supported by the ONR Contracts N00014-11-1-0712, N00014-12-1-0561, N00014-15-1-2181; the NSF Grants DMS 1222715, DMS 1222390; the Institut Universitaire de France; the ERC Adv grant BREAD; the DFG SFB-Transregio 40; the DFG Research Group 1779; the Excellence Initiative of the German Federal and State Governments, (RWTH Aachen Distinguished Professorship, Graduate School AICES); and the Polish NCN grant DEC2011/03/B/ST1/04902.

1 Introduction

1.1 Background and motivation

The emergence of computational and experimental engineering has led to a spectrum of new mathematical questions on how to best merge *data driven* and *model based* approaches. The development of corresponding *data-assimilation* methodologies has been originally driven mainly by meteorological research (see e.g. [12, 14]) but has meanwhile entered numerous areas in science and engineering bringing, in particular, the role of reduced order modeling into the focus of attention [1].

The present paper addresses some principal mathematical aspects that arise when trying to numerically capture a function u which is a state of a physical process with a known law, however with unknown parameters. We are given measurements of this state and the question is how to best merge these measurements with the model information to come up with a good approximation to u.

A typical setting of this type occurs when *all states* of the physical process are described by a specific *parametric family of PDEs* which is known to us, in a form

$$\mathcal{P}(u, \mu) = 0,$$

where μ is a vector of parameters ranging in a finite or infinite dimensional set $\mathcal{P}$. Instead of knowing the exact value of μ which would allow us to compute the state $u = u(\mu)$ by solving the equation, we observe one of these states through some collection of measurements and we want to use these measurements, together with the known parametric PDE, to numerically capture the state, or perhaps even more ambitiously to capture the parameters. Since the solution manifold

$$\mathcal{M} := \{u(\mu) \ : \ \mu \in \mathcal{P}\},$$

to a parametric PDE is generally quite complicated, it is usually seen through a sequence of nested finite dimensional spaces

$$V_0 \subset V_1 \subset \cdots \subset V_n, \quad \dim(V_j) = j,$$

such that each V_j approximates $\mathcal{M}$ to a known tolerance ε_j. Construction of such spaces is sometimes referred to as *model reduction*. Various algorithms for generating such spaces, together with error bounds ε_j, have been derived and analyzed. One of the most prominent of these is the *reduced basis method* where the spaces are generated through particular solution instances $u(\mu^i)$ picked from $\mathcal{M}$, see [5, 2, 11, 19]. Other algorithms with known error bounds are based on polynomial approximations in the parametric variable, see [7, 8].

Thus, the information that the state u we wish to approximate is on the manifold is replaced by the information of how well u can be approximated by the spaces V_j. Of course, this is not enough information to pin down u since we do not know where u is on the manifold, or in the new formulation, which particular element of V_j provides a good approximation to u. However, additional information about u is given by physical measurements which hopefully are enough to approximately locate u. This type of recovery problem was formulated and analyzed in [16] using an infinite dimensional Hilbert space setting which allows one to properly exploit the nature of the continuous background model when assimilating observations. This is also the setting adopted in the present paper.

The achievements of the present paper are two-fold. First, we establish that the algorithm proposed in [16] for estimating a state from a given set of observations and the knowledge of its approximability from a space V_n is *best possible in the sense of optimal recovery.* Second, and more importantly, we demonstrate the potential gain in accuracy for state recovery when combining the approximability by *each* of the subspaces V_j in the given hierarchy. We refer to this as the *multi-space setting* which will be seen to better exploit the information given by reduced bases or polynomial constructions. We give algorithms and performance bounds for these recovery algorithms in the multi-space setting when the observations are fixed and given to us. These algorithms are online implementable, similar to the ones discussed in [16]. Let us mention that one emphasis in [16] is on the selection of the measurement functionals in order to optimize the recovery process, while in the present paper we consider such functionals as given and focus on optimal recovery as explained above.

1.2 Conceptual preview

We study the above problems in the general framework of *optimal recovery* in a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. Under this setting, we are wanting to recover a function $u \in \mathcal{H}$ from its measurements $\ell_i(u) = \langle u, \omega_i \rangle$, where the ω_i are known elements of $\mathcal{H}$, $i = 1, \dots, m$. If we denote by W the space spanned by the ω_i, $i = 1, \dots, m$, then, the measurements determine $w = P_W u$ where throughout this paper P_X denotes the orthogonal projection onto X for any closed subspace $X \subset \mathcal{H}$. In going further, we think of measurements as simply providing the knowledge of this projection. In particular, we assume that the ω_j's are linearly independent i.e., $\dim W = m$. Therefore, our problem is to find an approximation $\hat{u}(w)$ to u from the information $w \in W$. This is equivalent to constructing a mapping $A : W \to \mathcal{H}$ and setting $\hat{u}(w) = A(w) = A(P_W u)$.

All elements of the orthogonal complement $W^\perp$ of W have zero measurements. A first observation is that if all the information we have about u is that $P_W u = w$, then we cannot recover u to any guaranteed accuracy. Indeed, if u_0 satisfies the measurements then u could be any of the functions $u_0 + \eta$, with $\eta \in W^\perp$, and each of these functions would be assigned the same approximation $\hat{u} = \hat{u}(w)$. Therefore, we need additional information about u to have a meaningful problem. A typical assumption is that u is in some known compact set $\mathcal{S} \subset \mathcal{H}$. The recovery problem in this case is known as *optimal recovery.* A classical setting is that $\mathcal{H}$ is the space L_2 and $\mathcal{S}$ is a finite ball in a Sobolev or Besov space, see e.g. [3, 17, 18].

In contrast to the case where $\mathcal{S}$ is a known Sobolev or Besov ball, our interest is in the setting where $\mathcal{S}$ is the solution manifold $\mathcal{M}$ of a parametric PDE. As noted above, the typical way of resolving $\mathcal{M}$ is through a finite sequence of spaces $\{V_0, \dots, V_n\}$ with V_k of dimension k where the spaces are known to approximate $\mathcal{M}$ to some known accuracy. This leads us to the following two settings:

The one-space problem: We assume that all what we know about $\mathcal{M}$ is that there is a space V_n of dimension n which is an approximation to $\mathcal{M}$ with accuracy ε_n. Accordingly, we define

$$\mathcal{K} := \mathcal{K}^{\text{one}} := \{u \in \mathcal{H} : \operatorname{dist}(u, V_n) \le \varepsilon_n\}, \tag{1.1}$$

and consider $u \in \mathcal{K}$ to be the only information we have about $\mathcal{M}$. In this case, the information (1.1) is the additional knowledge we have about u. We want to combine this knowledge with our measurements $P_W u$ to construct a good approximation $\hat{u}$ to u. So in this case, the spaces V_n and W are known and fixed.

The multi-space problem: We assume that what we know about $\mathcal{M}$ is that there is a sequence of spaces $V_0 \subset V_1 \subset \cdots \subset V_n$ such that each V_k has dimension k and approximates $\mathcal{M}$ with accuracy ε_k, where $\varepsilon_0 \geq \varepsilon_1 \geq \cdots \varepsilon_n > 0$. This leads us to define

$$\mathcal{K} := \mathcal{K}^{\text{mult}} := \bigcap_{j=0}^{n} \mathcal{K}^j, \tag{1.2}$$

where

$$\mathcal{K}^j := \{u \in \mathcal{H} : \operatorname{dist}(u, V_j) \leq \varepsilon_j\}, \quad j = 0, \dots, n.$$

In this case, the information $u \in \mathcal{K}$ is the additional knowledge we have about u. We want to combine this knowledge with our measurements to construct a good approximation $\hat{u}$ to u. As already noted, the multi-space problem is typical when applying reduced bases or polynomial methods to parametric PDEs.

1.3 Performance criteria

This paper is concerned with approximating a function $u \in \mathcal{H}$ from the information that $u \in \mathcal{K}$ and $P_W u = w$ in the two above settings. Note that in both settings, the set $\mathcal{K}$ is not compact. The additional information provided by the measurements gives that u is in the class

$$\mathcal{K}_w := \{u \in \mathcal{K} : P_W u = w\}.$$

This set is the intersection of $\mathcal{K}$ with the affine space

$$\mathcal{H}_w := \{u \in \mathcal{H} : P_W u = w\} = w + W^{\perp}.$$

Note that $\mathcal{K}_w$ may be an empty set for certain $w \in W$.

Recall that an algorithm is a mapping $A : W \to \mathcal{H}$ which assigns to any $w \in W$ the approximation $\hat{u}(w) = A(P_w u)$. In designing an algorithm, we are given the information of the spaces $(V_k)_{k=0,\dots,n}$ and the error bounds $(\varepsilon_k)_{k=0,\dots,n}$. There are several ways in which we can measure the performance of an algorithm. Consider first the one-space problem. A first way of measuring the performance of an algorithm is to ask for an estimate of the form

$$\|u - A(P_W u)\| \leq C_A(w) \operatorname{dist}(u, V_n), \quad u \in \mathcal{K}_w. \tag{1.3}$$

The best algorithm A, for a given fixed value of w, would give the smallest constant $C_A(w)$ and the algorithm which gives this smallest constant is said to be *instance optimal* with constant $C_A(w)$. In this case, the performance bound given by the right side of (1.3) depends not only on w but on the particular u from $\mathcal{K}_w$.

The estimate (1.3) also gives a performance bound for the entire class $\mathcal{K}_w$ in the form

$$\sup_{u \in \mathcal{K}_w} \|u - A(P_W u)\| \leq C_A(w) \varepsilon_n.$$

This leads us to the notion of performance of a recovery algorithm A on any set $\mathcal{S} \subset \mathcal{H}$ which is defined by

$$E_A(\mathcal{S}) := \sup_{u \in \mathcal{S}} \|u - A(P_W u)\|.$$

The *class optimal performance* on the set $\mathcal{S}$ is given by

$$E(\mathcal{S}) := \inf_{A} E_A(\mathcal{S}), \tag{1.4}$$

where the infimum is taken over all possible algorithms, i.e., all maps $A : W \to \mathcal{H}$. In particular, class optimal performance is defined for both the single space or multi-space settings and for both the sets $\mathcal{K}_w$ for each of individual w which gives the measure $E(\mathcal{K}_w)$ or the entire class $\mathcal{K}$ which gives the performance $E(\mathcal{K})$. The latter notion is the most meaningful when in applications it is not known which measurements $w \in W$ will appear or will be available.

The present paper studies each of the above problems with the goal of determining the best algorithms. For this purpose, we introduce for any closed subspaces V and W of $\mathcal{H}$ the quantity

$$\mu(V, W) := \sup_{\eta \in W^\perp} \frac{\|\eta\|}{\|\eta - P_V \eta\|} = \sup_{\eta \in W^\perp} \frac{\|\eta\|}{\|P_{V^\perp} \eta\|}. \tag{1.5}$$

A simple calculation shows that $\mu(V, W) = \beta(V, W)^{-1}$ where

$$\beta(V, W) := \inf_{v \in V} \frac{\|P_W v\|}{\|v\|} = \inf_{v \in V} \sup_{w \in W} \frac{\langle v, w \rangle}{\|v\| \|w\|}.$$

Note that in the case where $V = \{0\}$ we have $\mu(V, W) = 1$.

In §2 of the paper, we analyze the one space problem, that is, $\mathcal{K} = \mathcal{K}^{\rm one}$. The inf-sup constant β was used in [16], for the study of this problem, where the authors proposed an algorithm, in the form of a certain linear mapping $A^* : w \to A^*(w)$, then analyze its performance. While the approach in [16] is based on variational arguments, ours is quite different and geometric in nature. Our first goal is to establish that the algorithm proposed in [16] is both instance optimal and class optimal. We show that for any function $u \in \mathcal{H}$

$$\|u - A^*(P_W u)\| \le \mu(V_n, W) \operatorname{dist}(u, V_n). \tag{1.6}$$

Notice that if $\beta(V_n, W) = 0$, the above estimate would give no bound on approximation as is to be expected since V_n would contain elements of $W^\perp$ and these cannot be distinguished by the measurements. This would always be the case if $n > m$ and so in going further we always work under the assumption that $n \le m$.

Let us note that this is a modest improvement on the estimate in [16] which has the constant $\mu(V_n, W) + 1$ rather than $\mu(V_n, W)$ on the right side of (1.6). More importantly, we show that the estimate (1.6) is best possible in the sense that the constant $\mu(V_n, W)$ cannot be replaced by a smaller constant. Another important remark, observed in [16], is that in (1.6), $\operatorname{dist}(u, V_n)$ can be replaced by the smaller quantity $\operatorname{dist}(u, V_n \oplus (W \cap V_n^\perp))$. We establish, with our approach, the estimate

$$\|u - A^*(P_W u)\| \le \mu(V_n, W) \operatorname{dist}(u, V_n \oplus (W \cap V_n^\perp)), \tag{1.7}$$

which improves the constant given in [16]. We again show that $\mu(V_n, W)$ is the best constant in estimates of this form.

In view of (1.6), the algorithm A^* provides the class estimate

$$E_{A^*}(\mathcal{K}) \le \mu(V_n, W) \varepsilon_n. \tag{1.8}$$

We again show that this algorithm is *class optimal* in the sense that for the single space problem

$$E(\mathcal{K}) = \mu(V_n, W)\varepsilon_n.$$

Our analysis is based on proving lower bounds which show that the upper estimates (1.7) and (1.8) cannot be improved. These lower bounds apply to both linear and nonlinear algorithms, that is, (1.7) and (1.8) cannot be improved also using nonlinear mappings.

Another goal of our analysis of the one-space problem is to simplify the description of the optimal solution through the choice of, what we call, *favorable bases* for the spaces V_n and W. These favorable bases are then used in our analysis of the multi-space problem which is the object of §3. One possible way of proceeding, in the multi-space case, is to examine the right side of (1.8) for each of the spaces $(V_k)_{k=0,\dots,n}$, and choose the one which gives the minimum value. This would produce an algorithm A with the error bound

$$E_A(\mathcal{K}) \leq \min_{0\leq k\leq n} \mu(V_k, W)\varepsilon_k. \tag{1.9}$$

Notice that the ε_k are decreasing but the $\mu(V_k, W)$ are increasing as k gets larger. So these two quantities are working against one another and the minimum may be assumed for an intermediate value of k.

It turns out that the algorithm giving the bound (1.9) may be far from optimal and our main achievements in §3 are to produce both algorithms and a priori performance bounds which in general are better than that of (1.9). We show how the multi-space problem is connected to finding a point in the intersection of a family of ellipsoids in $\mathcal{H}$ and propose an algorithm based on this intersection property. Then, we give *a priori bounds* on the performance of our numerical algorithm, which are shown to be, in general, better than (1.9).

2 The one-space problem

2.1 Preliminary remarks

We begin with some general remarks which can be applied to our specific problem. If $\mathcal{S} \subset \mathcal{H}$ is a bounded set and we wish to simultaneously approximate all of the elements in $\mathcal{S}$, then the best approximation is described by the center of the *Chebyshev ball* of $\mathcal{S}$, which is defined as the smallest closed ball that contains $\mathcal{S}$. To describe this ball, we first define the *Chebyshev radius*

$$\mathrm{rad}(\mathcal{S}) := \inf\{r : \mathcal{S} \subset B(v, r) \text{ for some } v \in \mathcal{H}\}.$$

The following well known lemma says that the Chebyshev ball exists and is unique.

Lemma 2.1 *If $\mathcal{S}$ is any bounded set in $\mathcal{H}$ with $R := \mathrm{rad}(\mathcal{S})$, then there exists a unique $v^* \in \mathcal{H}$ such that*

$$\mathcal{S} \subset B(v^*, R). \tag{2.1}$$

Proof: For any $v \in \mathcal{H}$, we define

$$R_{\mathcal{S}}(v) := \inf\{r : \mathcal{S} \subset B(v, r)\},$$

which is a well-defined function from $\mathcal{H}$ to $\mathbb{R}$. It follows from triangle inequality that $R_{\mathcal{S}} : \mathcal{H} \to \mathbb{R}$ is continuous. It is also easily seen that

$$\mathcal{S} \subset B(v, R_{\mathcal{S}}(v)).$$

By definition, $\mathrm{rad}(\mathcal{S}) = \inf_{v\in\mathcal{H}} R_{\mathcal{S}}(v)$. Now, consider any infimizing sequence $(v_j)_{j\in\mathbb{N}}$, i.e.,

$$\lim_{j\to\infty} R_{\mathcal{S}}(v_j) = \mathrm{rad}(\mathcal{S}).$$

We claim that $(v_j)_{j\in\mathbb{N}}$ is a Cauchy sequence. To see this, define $r_j := R_{\mathcal{S}}(v_j)$. For any fixed j and k and any $z \in \mathcal{S}$ we define $d_j := v_j - z$ and $d_k := v_k - z$. Then, $\|d_j\| \le r_j$, and $\|d_k\| \le r_k$. Therefore,

$$\begin{aligned}
\|v_j - v_k\|^2 &= \|d_j - d_k\|^2 = \langle d_j - d_k, d_j - d_k\rangle \\
&= 2\langle d_j, d_j\rangle + 2\langle d_k, d_k\rangle - \langle d_j + d_k, d_j + d_k\rangle \\
&= 2\|d_j\|^2 + 2\|d_k\|^2 - 4\Big\|\frac{1}{2}(d_j + d_k)\Big\|^2 \\
&\le 2r_j^2 + 2r_k^2 - 4\Big\|\frac{1}{2}(v_j + v_k) - z\Big\|^2.
\end{aligned}$$

Since $z \in \mathcal{S}$ is arbitrary we get

$$\|v_j - v_k\|^2 \le 2r_j^2 + 2r_k^2 - 4\Big[R_{\mathcal{S}}\Big(\frac{1}{2}(v_j + v_k)\Big)\Big]^2 \le 2r_j^2 + 2r_k^2 - 4\,\mathrm{rad}(\mathcal{S})^2.$$

Since $r_j, r_k \to \mathrm{rad}(\mathcal{S})$, this shows that $(v_j)_{j\in\mathbb{N}}$ is a Cauchy sequence and has a limit v^*, which by the continuity of $v \mapsto R_{\mathcal{S}}(v)$ satisfies $R_{\mathcal{S}}(v^*) = \mathrm{rad}(\mathcal{S})$. The uniqueness of v^* also follows from the above inequality by contradiction. By using the continuity of $v \mapsto R_{\mathcal{S}}(v)$ one easily shows that (2.1) holds. $\square$

We sometimes say that v^* in the above lemma is the *center* of $\mathcal{S}$. For any bounded set $\mathcal{S}$, the diameter of $\mathcal{S}$ is related to its Chebyshev radius $\mathrm{rad}(\mathcal{S})$ by the inequalities

$$\mathrm{rad}(\mathcal{S}) \le \mathrm{diam}(\mathcal{S}) \le 2\,\mathrm{rad}(\mathcal{S}).$$

For general sets $\mathcal{S}$ these inequalities cannot be improved. However, we have the following remark.

Remark 2.2 *Let $\mathcal{S}$ be symmetric about a point z, i.e. whenever $v \in \mathcal{S}$, then $2z - v \in \mathcal{S}$. Then, the Chebyshev radius of $\mathcal{S}$ equals half its diameter, that is,* $\mathrm{diam}(\mathcal{S}) = 2\,\mathrm{rad}(\mathcal{S})$ *and its center is z.*

Remark 2.3 *In the particular setting of this paper, for any given $w \in W$ such that $\mathcal{K}_w$ is non-empty, the optimal recovery $u^*(w)$ over the class $\mathcal{K}_w$ is obviously given by the center of $\mathcal{K}_w$, and the class optimal performance is given by*

$$E(\mathcal{K}_w) = \mathrm{rad}(\mathcal{K}_w).$$

Remark 2.4 *For a bounded, closed, convex set $\mathcal{S} \subset \mathcal{H}$ (which is always the case in this paper) its center u is in $\mathcal{S}$. In fact, if this was not true, by translating $\mathcal{S}$, we can assume $u = 0$. Let $s_0 = \mathrm{argmin}_{s\in\mathcal{S}} \|s\|$. By convexity s_0 exists, $s_0 \ne 0$, and $\langle s, s_0\rangle \ge \langle s_0, s_0\rangle$, $s \in \mathcal{S}$. Thus*

$$\sup_{s\in\mathcal{S}} \|s - s_0\|^2 = \sup_{s\in\mathcal{S}}(\langle s, s\rangle - 2\langle s, s_0\rangle + \langle s_0, s_0\rangle) \le \sup_{s\in\mathcal{S}} \|s\|^2 - \|s_0\|^2$$

which contradicts the assumption that 0 is the center of $\mathcal{S}$.

2.2 Optimal bounds for the one-space problem

We next consider the case where the set $\mathcal{K} = \mathcal{K}^{\rm one}$ is given by (1.1), where V_n is a fixed and known n dimensional space. In this section, we derive the algorithm proposed in [16], however from a different point of view emphasizing more the optimal recovery and geometric aspects of the problem. This allows us to improve on their estimates some but, more importantly, it is also useful when treating the multi-space problem.

In the event that $\beta(V_n, W) = 0$, the space V_n contains elements from $W^\perp$ which implies that if $w \in W$ is such that $\mathcal{K}_w$ is non-empty, then $\mathcal{K}_w$ is unbounded, or equivalently $\mathrm{rad}(\mathcal{K}_w)$ *is infinite, which means that we cannot hope for any guaranteed performance over $\mathcal{K}_w$. This is the case in particular when $n > m$. For this reason, in the rest of the paper, we always assume that $\beta(V_n, W) > 0$, which means in particular that $n \le m$.*

Let w be any element from W. We claim that the map

$$u \mapsto \|u - P_{V_n} u\| = \|P_{V_n^\perp} u\|,$$

admits a unique minimizer over the affine space $\mathcal{H}_w$. To see this, we let u_0 be any element from $\mathcal{H}_w$. It follows that every $u \in \mathcal{H}_w$ can be written as $u = u_0 + \eta$ for some $\eta \in W^\perp$. Minimizing $\|P_{V_n^\perp} u\|$ over $\mathcal{H}_w$ therefore amounts to minimizing the function

$$\eta \mapsto f(\eta) := \|P_{V_n^\perp} u_0 + P_{V_n^\perp} \eta\|^2,$$

over $W^\perp$. We may write

$$f(\eta) := g(\eta) + \|P_{V_n^\perp} \eta\|^2,$$

where g is an affine function. Since we have assumed that $\beta(V_n, W) > 0$, the inequalities

$$\beta(V_n, W)\|\eta\| \le \|P_{V_n^\perp} \eta\| \le \|\eta\|, \quad \eta \in W^\perp.$$

show that $\eta \mapsto \|P_{V_n^\perp} \eta\|$ is an equivalent norm over $W^\perp$. Therefore $\eta \mapsto f(\eta)$ is strongly convex over $W^\perp$ and therefore admits a unique minimizer

$$\eta^* := \operatorname*{argmin}_{\eta \in W^\perp} f(\eta).$$

It follows that $u^* = u_0 + \eta^*$ satisfies

$$u^* = u^*(w) := \operatorname*{argmin}_{u \in \mathcal{H}_w} \|u - P_{V_n} u\|$$

and that this minimizer is unique.

Remark 2.5 *If w is such that $\mathcal{K}_w$ is non-empty, there exists a $u \in \mathcal{H}_w$ such that $\|u - P_{V_n} u\| \le \varepsilon_n$. Therefore $\|u^* - P_{V_n} u^*\| \le \varepsilon_n$, that is, $u^* \in \mathcal{K}_w$. In particular, u^* minimizes $\|u - P_{V_n} u\|$ over all $u \in \mathcal{K}_w$.*

We next define

$$v^* := v^*(w) := P_{V_n} u^*.$$

From the definition of u^*, it follows that the pair (u^*, v^*) is characterized by the minimization property

$$\|u^* - v^*\| = \min_{u \in \mathcal{H}_w,\, v \in V_n} \|u - v\|, \tag{2.2}$$

As the following remark shows, $u^* - v^*$ has a certain double orthogonality property.

Remark 2.6 *The element $u^* - v^*$ is orthogonal to both spaces V_n and $W^\perp$. The orthogonality to V_n follows from the fact that $v^* = P_{V_n} u^*$. On the other hand, for any $\eta \in W^\perp$ and $\alpha \in \mathbb{R}$, we have*

$$\|u^* - v^*\|^2 \le \|u - P_{V_n} u\|^2, \quad u := u^* + \alpha\eta,$$

and thus

$$\|u^* - v^*\|^2 \le \|u^* - v^* + \alpha(\eta - P_{V_n}\eta)\|^2 = \|u^* - v^*\|^2 + 2\alpha\langle u^* - v^*, \eta\rangle + \alpha^2\|\eta - P_{V_n}\eta\|^2.$$

This shows that $u^* - v^*$ is orthogonal to $W^\perp$.

Remark 2.7 *Conversely, if $u \in \mathcal{H}_w$ and $v \in V_n$ are such that $u - v$ is orthogonal to both spaces V_n and $W^\perp$, then $u = u^*$ and $v = v^*$. Indeed, from this orthogonality*

$$\|u^* - v^*\|^2 = \|u - v\|^2 + \|u^* - v^* - (u - v)\|^2.$$

This gives that u, v is also a minimizing pair and from uniqueness of the minimizing pair $u = u^*$ and $v = v^*$.

The next theorem describes the smallest ball that contains $\mathcal{K}_w$, i.e., the Chebyshev ball for this set, and shows that the center of this ball is $u^*(w)$.

Theorem 2.8 *Let W and V_n be such that $\beta(V_n, W) > 0$.*
(i) *For any $w \in W$ such that $\mathcal{K}_w$ is non-empty, the Chebyshev ball for $\mathcal{K}_w$ is the ball centered at $u^*(w)$ of radius*

$$R^* = R^*(w) := \mu(V_n, W)(\varepsilon_n^2 - \|u^*(w) - v^*(w)\|^2)^{1/2}. \tag{2.3}$$

(ii) *The optimal algorithm in the sense of* (1.4) *for recovering $\mathcal{K}_w$ from the measurement w is given by the mapping $A^* : w \mapsto u^*(w)$ and gives the performance bound*

$$E_{A^*}(\mathcal{K}_w) = E(\mathcal{K}_w) = \mu(V_n, W)(\varepsilon_n^2 - \|u^*(w) - v^*(w)\|^2)^{1/2}. \tag{2.4}$$

(iii) *The optimal algorithm in the sense of* (1.4) *for recovering $\mathcal{K}$ is given by the mapping $A^* : w \mapsto u^*(w)$ and gives the performance bound*

$$E_{A^*}(\mathcal{K}) = E(\mathcal{K}) = \mu(V_n, W)\varepsilon_n. \tag{2.5}$$

Proof: In order for $\mathcal{K}_w$ to be nonempty, we need that $\|u^* - v^*\| \le \varepsilon_n$. Any $u \in \mathcal{H}_w$ can be written as $u = u^* + \eta$ where $\eta \in W^\perp$. Therefore,

$$u - P_{V_n} u = u^* - v^* + \eta - P_{V_n}\eta.$$

Because of the orthogonality in Remark 2.6, we have

$$\|u - P_{V_n} u\|^2 = \|u^* - v^*\|^2 + \|\eta - P_{V_n}\eta\|^2. \tag{2.6}$$

Thus a necessary and sufficient condition for u to be in $\mathcal{K}_w$ is that

$$\|P_{V_n^\perp}\eta\|^2 = \|\eta - P_{V_n}\eta\|^2 \le \varepsilon_n^2 - \|u^* - v^*\|^2.$$

From the definition of $\mu(V_n, W)$, this means that any $u \in \mathcal{K}_w$ is contained in the ball $B(u^*(w), R^*(w))$. Now, if η is any element in $W^\perp$ with norm $R^*(w)$ which achieves the maximum in the definition of $\mu(V_n, W)$, then $u^* \pm \eta$ is in $\mathcal{K}_w$ and since $\|\eta\| = R^*(w)$ we see that the diameter of $\mathcal{K}_w$ is at least as large as $2R^*(w)$. Since $\mathcal{K}_w$ is the translation of a symmetric set, we thus obtain (i) from Remark 2.2. The claim (ii) about A^* being the optimal algorithm follows from Remark 2.3. Finally, the performance bound (2.5) in the claim (iii) holds because the maximum of $R^*(w)$ is achieved when $w = 0$. □

Remark 2.9 *The optimal mapping $w \mapsto A^*(w) = u^*(w)$ is independent of ε_n and the knowledge of ε_n is not needed in order to compute $A^*(w)$.*

Remark 2.10 *Since $\mathcal{K}_w$ is the intersection of the cylinder $\mathcal{K}$ with the affine space $\mathcal{H}_w$, it has the shape of an ellipsoid. The above analysis describes this ellipsoid as follows: a point $u^* + \eta$ is in $\mathcal{K}_w$ if and only if $\|P_{V_n^\perp}\eta\|^2 \le \varepsilon_n^2 - \|u^* - v^*\|^2$. In the following section, we give a parametric description of this ellipsoid using certain coordinate systems, see Lemma 2.14.*

Remark 2.11 *The elements u^* and v^* were introduced in [16] and used to define the algorithm A^* given in the above theorem. The analysis from [16] establishes the error bound*

$$\|u - u^*(w)\| \le (\mu(V_n, W) + 1)\,\mathrm{dist}(u, V_n \oplus (V_n^\perp \cap W)).$$

A sharper form of this inequality can be derived from our results. Namely, if u is any element in $\mathcal{H}$ then we can define $\varepsilon_n := \|u - P_{V_n} u\|$ and $w := P_W u$. Then, $u \in \mathcal{K}_w$, for this choice of ε_n, and so Theorem 2.8 applies and gives a recovery of u with the bound

$$\|u - u^*(w)\| \le \mu(V_n, W)(\varepsilon_n^2 - \|u^* - v^*\|^2)^{1/2} = \mu(V_n, W)\|u - P_{V_n} u - (u^* - v^*)\|, \tag{2.7}$$

where the second equality follows from (2.6). We have noticed in Remark 2.6 that $u^* - v^* \in V_n^\perp \cap W$, and on the other hand we have that $u - (u^* - v^*) \in V_n + W^\perp$, which shows that

$$u^* - v^* = P_{V_n^\perp \cap W} u.$$

Therefore

$$P_{V_n} u + u^* - v^* = P_{V_n \oplus (V_n^\perp \cap W)} u,$$

and (2.7) gives

$$\|u - u^*(w)\| \le \mu(V_n, W)\,\mathrm{dist}(u, V_n \oplus (V_n^\perp \cap W)).$$

Remark 2.12 *Let us observe that given a space V_n with $n < m$ we have $(W \cap V_n^\perp) \neq \{0\}$, thus the space $\bar{V}_n := V_n \oplus (W \cap V_n^\perp)$ is strictly larger than V_n. However $\mu(\bar{V}_n, W) = \mu(V_n, W)$ because for any $\eta \in W^\perp$, the projection of η onto $W \cap V_n^\perp$ is zero. In other words we can enlarge V_n preserving the estimate* (2.5) *for class optimality performance as long as we add parts of W that are orthogonal to V_n.*

2.3 The numerical implementation of the optimal algorithm

Let us next discuss the numerical implementation of the optimal algorithm for the one-space problem. Let $\omega_1, \dots, \omega_m$ be any orthonormal basis for W. For theoretical reasons only, we complete it to an orthonormal basis for $\mathcal{H}$. So $\{\omega_i\}_{i>m}$ is a complete orthonormal system for $W^\perp$. We can write down explicit formulas for u^* and v^*. Indeed, any $u \in \mathcal{H}_w$ can be written

$$u = \sum_{i=1}^{m} w_i \omega_i + \sum_{i=m+1}^{\infty} x_i \omega_i,$$

where $w_i := \langle w, \omega_i \rangle$, and $(x_i)_{i>m}$ is any ℓ_2 sequence. So, for any $v \in V_n$ and $u \in \mathcal{H}_w$, we have

$$\|u - v\|^2 = \sum_{i=1}^{m} (w_i - v_i)^2 + \sum_{i=m+1}^{\infty} (x_i - v_i)^2,$$

where $v_i := \langle v, \omega_i \rangle$. Thus, for any $v \in V_n$, its best approximation $u(v)$ from $\mathcal{H}_w$ is

$$u(v) := \sum_{i=1}^{m} w_i \omega_i + \sum_{i=m+1}^{\infty} v_i \omega_i, \tag{2.8}$$

and its error of approximation is

$$\|v - u(v)\|^2 = \sum_{i=1}^{m} (w_i - v_i)^2.$$

In view of (2.2) we have

$$v^* = \operatorname*{argmin}_{v \in V_n} \|v - u(v)\|^2 = \operatorname*{argmin}_{v \in V_n} \sum_{i=1}^{m} (w_i - v_i)^2 = \operatorname*{argmin}_{v \in V_n} \|w - P_W v\|^2.$$

For any given orthonormal basis $\{\phi_1, \cdots, \phi_n\}$ for V_n, we can find the coordinates of $v^* \in V_n$ in this basis by solving the $n \times n$ linear system associated to the above least squares problem. Once v^* is found, the optimal recovery $u^* = u^*(w)$ is given, according to (2.8), by

$$u^* = v^* + \sum_{i=1}^{m} (w_i - v_i^*) \omega_i,$$

where $v_i^* = \langle v^*, \omega_i \rangle$. Note that we may also write

$$u^* = \sum_{i=1}^{m} w_i \omega_i + \sum_{i=m+1}^{\infty} \langle v^*, \omega_i \rangle \omega_i = w + P_{W^\perp} v^*. \tag{2.9}$$

2.4 Liftings and favorable bases for V_n and W

It turns out that the above optimal algorithm has an even simpler description if we choose suitable bases for V_n and W, which we call *favorable bases.* These bases will also be important in our analysis of the multi-space problem. To describe this new geometric view, we introduce the description of algorithms through liftings and see how the best algorithm of the previous section arises in this context.

As noted earlier, any algorithm is a mapping $A : W \to \mathcal{H}$ which takes $w = P_W u$ into $\hat{u}(w) = A(w) = A(P_W u)$. This image serves as the approximant of all of the $u \in \mathcal{K}_w$. We can write any $u \in \mathcal{K}_w$ as $u = w + P_{W^\perp} u$. So the problem is to find an appropriate mapping $F : W \to W^\perp$ and take as the approximation

$$\hat{u}(w) := A(w) := w + F(w).$$

At this stage F can be any linear or nonlinear mapping from W into $W^\perp$. We call such mappings F *liftings.*

According to (2.9), the optimal lifting F^* is defined by

$$F^*(w) = P_{W^\perp} v^*(w) \in P_{W^\perp} V_n,$$

which is actually a linear mapping since v^* depends linearly on w. The algorithm $A^*(w) = w + F^*(w)$ was shown in the previous section to be optimal for each class $\mathcal{K}_w$ as well as for $\mathcal{K}$. Note that this optimality holds even if we open the competition to nonlinear maps F, respectively A.

We next show that F^* has a simple description as a linear mapping by introducing favorable bases. We shall make use of the following elementary facts from linear algebra: if X and Y are closed subspaces of a Hilbert space $\mathcal{H}$, then:

- We have the equality

$$\dim(P_X Y) = \dim(P_Y X).$$

  This can be seen by introducing the cross-Gramian matrix $G = (\langle x_i, y_j \rangle)$, where (x_i) and (y_j) are orthonormal bases for X and Y. Then G is the matrix representation of the projection operator P_X from Y onto X with respect to these bases and G^t is the corresponding representation of the projection operator P_Y from X onto Y. Hence,

$$\dim(P_X Y) = \operatorname{rank}(G) = \operatorname{rank}(G^t) = \dim(P_Y X).$$

- The space Y can be decomposed into a direct orthogonal sum

$$Y = P_Y X \oplus (Y \cap X^\perp). \tag{2.10}$$

  For this, we need to show that $Y \cap X^\perp = Z$ where $Z \subset Y$ is the orthogonal complement of $P_Y X$ in Y. If $y \in Z$, then $\langle y, P_Y x \rangle = 0$ for all $x \in X$. Since $\langle y, x - P_Y x \rangle = 0$, if follows that $\langle y, x \rangle = 0$, for all $x \in X$, and thus $y \in Y \cap X^\perp$. Conversely if $y \in Y \cap X^\perp$, then for any $x \in X$ $\langle y, P_Y x \rangle = -\langle y, x - P_Y x \rangle = 0$, which shows that $y \in Z$.

Now to construct the favorable bases we want, we begin with any orthonormal basis $\{\phi_1, \dots, \phi_n\}$ of V_n and any orthonormal basis $\{\omega_1, \dots, \omega_m\}$ of W. We consider the $m \times n$ cross-Gramian matrix

$$G := (\langle \omega_i, \phi_j \rangle),$$

which may be viewed as the matrix representation of the projection operator P_W from V_n onto W using these bases since $P_W(\phi_j) = \sum_{i=1}^m \langle \omega_i, \phi_j \rangle \omega_i$. Note that the inf-sup condition $\beta(V_n, W) > 0$ means that

$$\dim(P_W V_n) = n,$$

or equivalently, the rank of G is equal to n. We perform a singular value decomposition of G, which gives

$$G = USV^t$$

where $U = (u_{i,j})$ and $V = (v_{i,j})$ are unitary $m \times m$ and $n \times n$ matrices, respectively, and where S is an $m \times n$ matrix with entries $s_i > 0$ on the diagonal $i = j$, $i = 1, \dots, n$, and zero entries elsewhere. This allows us to define new orthonormal bases $\{\phi_1^*, \dots, \phi_n^*\}$ for V_n and $\{\omega_1^*, \dots, \omega_m^*\}$ for W by

$$\phi_j^* = \sum_{i=1}^n v_{i,j} \phi_i \quad \text{and} \quad \omega_j^* = \sum_{i=1}^m u_{i,j} \omega_i.$$

These new bases are such that

$$P_W(\phi_j^*) = s_j \omega_j^*, \quad j = 1, \dots, n,$$

and have diagonal cross-Gramian, namely

$$\langle \omega_i^*, \phi_j^* \rangle = s_j \delta_{i,j}.$$

Therefore $\{\omega_1^*, \dots, \omega_n^*\}$ and $\{\omega_{n+1}^*, \dots, \omega_m^*\}$ are orthonormal bases for the n-dimensional space $P_W V_n$ and respectively its orthogonal complement in W which is $V_n^\perp \cap W$ according to (2.10).

By convention, we organize the singular values in decreasing order

$$0 < s_n \le s_{n-1} \le \cdots \le s_1.$$

Since P_W is an orthogonal projector, all of them are at most 1 and in the event where

$$s_1 = s_2 = \cdots = s_p = 1,$$

for some $0 < p \le n$, then we must have

$$\omega_j^* = \phi_j^*, \quad j = 1, \dots, p.$$

This corresponds to the case where $V_n \cap W$ is non-trivial and $\{\omega_1^*, \dots, \omega_p^*\}$ forms an orthonormal basis for $V_n \cap W$. We define $p = 0$ in the case where $V_n \cap W = \{0\}$.

We may now give a simple description of the optimal algorithm A^* and lifting F^*, in terms of their action on the basis elements ω_j^*. For $j = n+1, \dots, m$, we know that $\omega_j^* \in V_n^\perp \cap W$. From Remark 2.7, it follows that the optimal pair (u^*, v^*) which solves (2.2) for $w = \omega_j^*$ is

$$u^* = \omega_j^* \quad \text{and} \quad v^* = 0,$$

and therefore

$$A^*(\omega_j^*) = \omega_j^* \quad \text{and} \quad F^*(\omega_j^*) = 0, \quad j = n+1, \dots, m.$$

For $j = 1, \dots, n$, we know that $\omega_j^* = P_W(s_j^{-1}\phi_j^*)$. It follows that the optimal pair (u^*, v^*) which solves (2.2) for $w = \omega_j^*$ is

$$u^* = v^* = s_j^{-1}\phi_j^*.$$

Indeed, this follows from Remark 2.7 since this pair has $u^* - v^* = 0$ and hence has the double orthogonality property. So, in this case,

$$A^*(\omega_j^*) = s_j^{-1}\phi_j^* \quad \text{and} \quad F^*(\omega_j^*) = s_j^{-1}\phi_j^* - \omega_j^*.$$

Note in particular that $F^*(\omega_j^*) = 0$ for $j = 1, \dots, p$.

Remark 2.13 *The favorable bases are useful when computing the inf-sup constant $\beta(V_n, W)$. Namely, for an element $v = \sum_{j=1}^n v_j\phi_j^* \in V_n$ we find that $P_W v = \sum_{j=1}^n s_j v_j \omega_j^*$ and so*

$$\beta(V_n, W) = \min_{v \in V_n} \frac{\|P_W v\|}{\|v\|} = \min_{v \in V_n} \left(\frac{\sum_{j=1}^n s_j^2 v_j^2}{\sum_{j=1}^n v_j^2} \right)^{1/2} = \min_{j=1,\dots,n} s_j = s_n.$$

Correspondingly,

$$\mu(V_n, W) = s_n^{-1}.$$

Recall that for the trivial space $V_0 = \{0\}$, we have $\mu(V_0, W) = 1$.

For further purposes, we complete the favorable bases into orthonormal bases of $\mathcal{H}$ by constructing particular orthonormal bases for $V_n^\perp$ and $W^\perp$. According to (2.10) we may write these spaces as direct orthogonal sums

$$V_n^\perp = P_{V_n^\perp}(W) \oplus (V_n^\perp \cap W^\perp),$$

and

$$W^\perp = P_{W^\perp}(V_n) \oplus (V_n^\perp \cap W^\perp).$$

The second space $V_n^\perp \cap W^\perp$ in the above decompositions may be of infinite dimension and we consider an arbitrary orthonormal basis $(\psi_i^*)_{i \geq 1}$ for this space. For the first spaces in the above decompositions, we can build orthonormal bases from the already constructed favorable bases.

For the space $P_{V_n^\perp}(W)$ we first consider the functions

$$P_{V_n^\perp}\omega_i^*, \quad i = 1, \dots, m$$

These functions are 0 for $i = 1, \dots, p$ since $\omega_i^* \in V_n$ for these values of i. They are equal to ω_i^* for $i = n+1, \dots, m$ and to $\omega_i^* - s_i\phi_i^*$ for $i = p+1, \dots, n$, and these $m-p$ functions are non-zero pairwise orthogonal. Therefore an orthonormal basis of $P_{V_n^\perp}(W)$ is given by the normalized functions

$$(1 - s_i^2)^{-1/2}(\omega_i^* - s_i\phi_i^*), \quad i = p+1, \dots, n, \quad \text{and} \quad \omega_i^*, \quad i = n+1, \dots, m.$$

By a similar construction, we find that an orthonormal basis of $P_{W^\perp}(V_n)$ is given by the normalized functions

$$(1-s_i^2)^{-1/2}(\phi_i^* - s_i\omega_i^*), \quad i = p+1,\dots,n.$$

Therefore bases for $V_n^\perp$ and $W^\perp$ are defined as union of these bases with the basis $(\psi_i^*)_{i\geq 1}$ for $V_n^\perp \cap W^\perp$.

Finally, we close out this section, by giving a parametric description of the set $\mathcal{K}_w = \mathcal{K}_w(V_n)$ for the single space problem which shows in particular that this set is an ellipsoid.

Lemma 2.14 *Given a single space $V_n \subset \mathcal{H}$, the body*

$$\mathcal{K}_w := \mathcal{K}_w(V_n) := \mathcal{K}_w^{\rm one}(V_n) := \{u \in \mathcal{K}^{\rm one}(V_n) : \ P_W u = w\}$$

is a non-degenerate ellipsoid contained in the affine space $\mathcal{H}_w$.

Proof: Using the favorable bases for W and $W^\perp$, we can write any $u \in \mathcal{H}_w$ as

$$u = \sum_{j=1}^m w_j\omega_j^* + \sum_{j=p+1}^n x_j(1-s_j^2)^{-1/2}(\phi_j^* - s_j\omega_j^*) + \sum_{i\geq 1} y_j\psi_j^*,$$

where the $w_j = \langle w, \omega_j^*\rangle$ for $j = 1,\dots,m$, are given, and the x_j and y_j are the coordinates of $u - w$ in the favorable basis of $W^\perp$. We may now write

$$\begin{aligned} P_{V_n^\perp}u &= \sum_{j=1}^m w_j P_{V_n^\perp}\omega_j^* + \sum_{j=p+1}^n x_j(1-s_j^2)^{-1/2}P_{V_n^\perp}(\phi_j^* - s_j\omega_j^*) + \sum_{i\geq 1} y_j\psi_j^* \\ &= \sum_{j=p+1}^m w_j(\omega_j^* - s_j\phi_j^*) - \sum_{j=p+1}^n x_j(1-s_j^2)^{-1/2}s_j(\omega_j^* - s_j\phi_j^*) + \sum_{i\geq 1} y_j\psi_j^* \\ &= \sum_{j=n+1}^m w_j(\omega_j^* - s_j\phi_j^*) + \sum_{j=p+1}^n (w_j - x_js_j(1-s_j^2)^{-1/2})(\omega_j^* - s_j\phi_j^*) + \sum_{i\geq 1} y_j\psi_j^*. \end{aligned}$$

All terms in the last sum are pairwise orthogonal and therefore

$$\|P_{V_n^\perp}u\|^2 = \sum_{j=n+1}^m (1-s_j^2)w_j^2 + \sum_{j=p+1}^n (1-s_j^2)(w_j - x_js_j(1-s_j^2)^{-1/2})^2 + \sum_{j\geq 1} y_j^2.$$

Now $u \in \mathcal{K}_w$ if and only if $\|P_{V_n^\perp}u\|^2 \leq \varepsilon_n^2$, or equivalently

$$\sum_{j=p+1}^n s_j^2(x_j - a_j)^2 + \sum_{j\geq 1} y_j^2 \leq C, \tag{2.11}$$

with $C := \varepsilon_n^2 - \sum_{j=n+1}^m (1-s_j^2)w_j^2$ and $a_j := (1-s_j^2)^{1/2}s_j^{-1}w_j$ which is the equation of a non-degenerate ellipsoid in $\mathcal{H}_w$. □

Remark 2.15 *The above equation* (2.11) *directly shows that the radius of $\mathcal{K}_w$ is equal to $s_n^{-1}C^{1/2}$ which is an equivalent expression of* (2.3).

3 The multi-space problem

In this section, we consider the multi-space problem as described in the introduction. We are interested in the optimal recovery of the elements in the set $\mathcal{K} := \mathcal{K}^{\mathrm{mult}}$ as described by (1.2). For any given $w \in W$, we consider the set

$$\mathcal{K}_w := \mathcal{K}_w^{\mathrm{mult}} := \mathcal{K}^{\mathrm{mult}} \cap \mathcal{H}_w = \bigcap_{j=0}^{n} \mathcal{K}_w^j,$$

where

$$\mathcal{K}_w^j := \mathcal{K}^j \cap \mathcal{H}_w := \{u \in \mathcal{H}_w \ : \ \mathrm{dist}(u, V_j) \le \varepsilon_j\}.$$

In other words, $\mathcal{K}_w^j$ is the set in the one-space problem considered in the previous section. We have seen that $\mathcal{K}_w^j$ is an ellipsoid with known center $u_j^* = u_j^*(w)$ and known Chebyshev radius given by (2.3) with n replaced by j, and u^* and v^* replaced by u_j^* and v_j^* in that formula.

Thus, $\mathcal{K}_w$ is now the intersection of $n+1$ ellipsoids. The optimal algorithm A^*, for the recovery of $\mathcal{K}_w$, is the one that would find the center of the Chebyshev ball of this set and its performance would then be given by its Chebyshev radius. In contrast to the one-space problem, this center and radius do not have simple computable expressions. The first results of this section provide an a priori estimate of the Chebyshev radius in the multi-space setting by exploiting favorable bases. This a priori analysis illustrates when a gain in performance is guaranteed to occur, although the a priori estimates we provide may be pessimistic.

We then give examples which show that the Chebyshev radius in the multi-space case can be far smaller than the minimum of the Chebyshev radii of the $\mathcal{K}_w^j$ for $j = 0, \dots, n$. These examples are intended to illustrate that exploiting the multi-space case can be much more advantageous than simply executing the one-space algorithms and taking the one with best performance, see (2.4).

The latter part of this section proposes two simple algorithmic strategies, each of them converging to a point in $\mathcal{K}_w$. These algorithms thus produce a near optimal solution, in the sense that if A is the map corresponding to either one of them, we have

$$E_A(\mathcal{K}_w) \le 2E_{A^*}(\mathcal{K}_w) = 2E(\mathcal{K}_w), \quad w \in W, \tag{3.1}$$

and in particular

$$E_A(\mathcal{K}) \le 2E(\mathcal{K}). \tag{3.2}$$

Both of these algorithms are iterative and based on alternating projections. An a posteriori estimate for the distance between a given iterate and the intersection of the ellipsoids is given and used both, as a stopping criteria and to analyze the convergence rates of the algorithms.

3.1 A priori bounds for the radius of $\mathcal{K}_w$

In this section, we derive a priori bounds for $\mathrm{rad}(\mathcal{K}_w^{\mathrm{mult}})$. Although these bounds may overestimate $\mathrm{rad}(\mathcal{K}_w^{\mathrm{mult}})$, they allow us to show examples where the multi-space algorithm is significantly better than simply chosing one space and using the one-space algorithm. Recall that for the one-space problem, we observed that $\mathrm{rad}(\mathcal{K}_w^{\mathrm{one}})$ is largest when $w = 0$. The following results show that for the multi-space problem $\mathrm{rad}(\mathcal{K}_w^{\mathrm{mult}})$ is also controlled by $\mathrm{rad}(\mathcal{K}_0^{\mathrm{mult}})$, up to a multiplicative constant. Note that $\mathcal{K}_w^{\mathrm{mult}}$ is generally not a symmetric set, except for $w = 0$. In going further in this section $\mathcal{K}$ and $\mathcal{K}_w$ will refer to the multi-space sets.

Lemma 3.1 *For the multi-space problem, one has*

$$\mathrm{rad}(\mathcal{K}_w) \leq 2\,\mathrm{rad}(\mathcal{K}_0), \quad w \in W. \tag{3.3}$$

Therefore,

$$E(\mathcal{K}) \leq 2\,\mathrm{rad}(\mathcal{K}_0). \tag{3.4}$$

Proof: Fix $w \in W$ and let $\tilde{u} := \tilde{u}(w)$ be the center of the Chebyshev ball for $\mathcal{K}_w$ which by Remark 2.4, belongs to $\mathcal{K}_w$. For any $u \in \mathcal{K}_w$ we have $\eta := \frac{1}{2}(u - \tilde{u})$ is in $W^\perp$ and also

$$\mathrm{dist}(\eta, V_k) \leq \frac{1}{2}(\mathrm{dist}(u, V_k) + \mathrm{dist}(\tilde{u}, V_k)) \leq \varepsilon_k, \quad k = 0, 1, \dots, n.$$

Hence, $\eta \in \mathcal{K}_0$ which gives

$$\|u - \tilde{u}\| = 2\|\eta\| \leq 2\,\mathrm{rad}(\mathcal{K}_0),$$

where we have used the fact that, by Remark 2.2, the best Chebyshev ball for $\mathcal{K}_0$ is centered at 0. This proves (3.3). The estimate (3.4) follows from the definition of $E(\mathcal{K})$. $\square$

In view of the above Lemma 3.1, we concentrate on deriving a priori bounds for the radius of the set $\mathcal{K}_0$. We know that $\mathcal{K}_0$ is the intersection of the ellipsoids $\mathcal{K}_0^j$ for $j = 0, 1, \dots, n$, each of which is centered at zero. We also know that the Chebyshev ball for $\mathcal{K}_0^j$ is $B(0, \mathrm{rad}(\mathcal{K}_0^j)$ and we know from (2.4) that

$$\mathrm{rad}(\mathcal{K}_0^j) = \mu(V_j, W)\varepsilon_j, \quad j = 0, 1, \dots, n,$$

which is a computable quantity. This gives the obvious bound

$$\mathrm{rad}(\mathcal{K}_0) \leq \min_{0 \leq k \leq n} \mu(V_k, W)\varepsilon_k. \tag{3.5}$$

In the following, we show that we can improve on this bound considerably. Since $\mathcal{K}_0$ is symmetric around the origin, we have

$$\mathrm{rad}(\mathcal{K}_0) = \operatorname*{argmax}_{\eta \in \mathcal{K}_0} \|\eta\|.$$

So we are interested in bounding $\|\eta\|$ for each $\eta \in \mathcal{K}_0$.

Since the spaces V_j are nested, we can consider an orthonormal basis $\{\phi_1, \dots, \phi_n\}$ for V_n, for which, $\{\phi_1, \dots, \phi_j\}$ is an orthonormal basis for each of the V_j for $j = 1, \dots, n$. We will use the favorable bases constructed in the previous section in the case of the particular space V_n. Note that if $\{\phi_1^*, \dots, \phi_n^*\}$ is the favorable basis for V_n, we do not generally have that $\{\phi_1^*, \dots, \phi_j^*\}$ is a basis of V_j.

Let η be any element from $\mathcal{K}_0$. Since $\mathrm{dist}(\eta, V_n) \leq \varepsilon_n$, we may express η as

$$\eta = \sum_{j=1}^n \eta_j \phi_j^* + e = \sum_{j=1}^n \alpha_j \phi_j + e, \quad e \in V_n^\perp \text{ and } \|e\| \leq \varepsilon_n.$$

So,

$$\|\eta\|^2 = \sum_{j=1}^n \eta_j^2 + \|e\|^2 = \sum_{j=1}^n \alpha_j^2 + \|e\|^2.$$

The α_j and η_j are related by the equations

$$\sum_{j=1}^{n} \lambda_{i,j}\alpha_j = \eta_i, \quad i = 1, \dots, n,$$

where

$$\lambda_{i,j} := \langle \phi_j, \phi_i^* \rangle, \quad 1 \le i, j \le n.$$

The fact that $\operatorname{dist}(\eta, V_k) \le \varepsilon_k$ for $k = 0, \dots, n$ is expressed by the inequalities

$$\sum_{j=k+1}^{n} \alpha_j^2 + \|e\|^2 \le \varepsilon_k^2, \quad k = 0, \dots, n.$$

Since $\eta \in W^\perp$, we have that

$$0 = P_W \eta = \sum_{j=1}^{n} s_j \eta_j \omega_j^* + P_W e\,.$$

It follows that

$$\sum_{j=1}^{n} s_j^2 \eta_j^2 = \|P_W e\|^2 \le \|e\|^2 \le \varepsilon_n^2.$$

We now return to the representation of $\mathcal{K}_0$ in the ϕ_j coordinate system. We know that all α_j satisfy $|\alpha_j| \le \varepsilon_{j-1}$. This means that the coordinates $\{\alpha_1, \dots, \alpha_n\}$ of any point in $\mathcal{K}_0$ are in the n-dimensional rectangle

$$R = [-\varepsilon_0, \varepsilon_0] \times \cdots \times [-\varepsilon_{n-1}, \varepsilon_{n-1}].$$

It follows that each η_i satisfies the crude estimate

$$|\eta_i| \le \sum_{j=1}^{n} |\lambda_{i,j}||\alpha_j| \le \sum_{j=1}^{n} |\lambda_{i,j}|\varepsilon_{j-1} =: \theta_i \quad i = 1, \dots, n. \tag{3.6}$$

The numbers θ_i are computable. The bound (3.6) allows us to estimate

$$\operatorname{rad}(\mathcal{K}_0)^2 = \sup_{\eta \in \mathcal{K}_0} \|\eta\|^2 \le \varepsilon_n^2 + \sup \Big\{ \sum_{j=1}^{n} \eta_j^2 : \ |\eta_j| \le \theta_j \quad \text{and} \quad \sum_{j=1}^{n} s_j^2 \eta_j^2 \le \varepsilon_n^2 \Big\}$$

Since the s_j are non-increasing, the supremum on the right side takes the form

$$\delta \theta_k^2 + \sum_{j=k+1}^{n} \theta_j^2, \quad 0 < \delta \le 1,$$

where k is the largest integer such that

$$\sum_{j=k}^{n} s_j^2 \theta_j^2 \ge \varepsilon_n^2, \tag{3.7}$$

and δ is chosen so that

$$\delta s_k^2 \theta_k^2 + \sum_{j=k+1}^{n} s_j^2 \theta_j^2 = \varepsilon_n^2. \tag{3.8}$$

This gives us the following bound on the Chebyshev radius of $\mathcal{K}_0$.

$$\mathrm{rad}(\mathcal{K}_0)^2 \leq \varepsilon_n^2 + \delta\theta_k^2 + \sum_{j=k+1}^{n} \theta_j^2 := E_n^2. \tag{3.9}$$

Using this estimate together with Lemma 3.1, we have proven the following theorem.

Theorem 3.2 *For the multi-space problem, we have the following estimates for Chebyshev radii. For $\mathcal{K}_0$, we have*

$$\mathrm{rad}(\mathcal{K}_0) \leq E_n,$$

where $E_n := \left(\varepsilon_n^2 + \delta\theta_k^2 + \sum_{j=k+1}^{n} \theta_j^2\right)^{1/2}$. For any $w \in W$, we have

$$\mathrm{rad}(\mathcal{K}_w) \leq 2E_n.$$

For $\mathcal{K}$, we have the bound

$$\mathrm{rad}(\mathcal{K}) \leq 2E_n.$$

We next compare the bound in (3.9) with the one space bound

$$\mathrm{rad}(\mathcal{K}_0) \leq \mu(V_n, W)\varepsilon_n = s_n^{-1}\varepsilon_n,$$

which is obtained by considering only the approximation property of V_n and not exploiting the other spaces V_j, $j < n$, see (3.5). For this, we return to the definition of k from (3.7). We can write each term that appears in (3.8) as $\gamma_j \varepsilon_n^2$ where $\sum_{j=k}^{n} \gamma_j = 1$. In other words,

$$\theta_j^2 = \gamma_j s_j^{-2} \varepsilon_n^2, \quad k < j \leq n, \quad \theta_k^2 = \delta^{-1}\gamma_k s_k^{-2}\varepsilon_n^2.$$

Hence,

$$E_n^2 \leq \varepsilon_n^2 + s_n^{-2}\varepsilon_n^2 \leq 2s_n^{-2}\varepsilon_n^2,$$

which is at least as good as the old bound up to a multiplicative constant $\sqrt{2}$.

We finally observe that the bound E_n is obtained by using the entire sequence $\{V_0, \dots, V_n\}$. Similar bounds E_Γ are obtained when using a subsequence $\{V_j \,:\, j \in \Gamma\}$ for any $\Gamma \subset \{0, \dots, n\}$. This leads to the improved bound

$$\mathrm{rad}(\mathcal{K}_0) \leq \min\{E_\Gamma \,:\, \Gamma \subset \{0, \dots, n\}\}.$$

In particular defining $E_j = E_\Gamma$ for $\Gamma = \{0, \dots, j\}$ we find that

$$E_j^2 \leq 2\mu(V_j, W)^2 \varepsilon_j^2.$$

Therefore

$$E_n^* \leq \sqrt{2} \min_{j=0,\dots,n} \mu(V_j, W)\varepsilon_j,$$

which shows that the new estimate is as good as (3.5) up to the multiplicative constant $\sqrt{2}$.

3.2 Examples

One can easily find examples for which the Chebyshev radius of $\mathcal{K}_w$ is substantially smaller than the minimum of the Chebyshev radii of the $\mathcal{K}_w^j$, therefore giving higher potential accuracy in the multi-space approach. As a simple example to begin this discussion, consider the case where

$$\mathcal{H} = \mathbb{R}^2, \quad V_0 = \{0\}, \quad V_1 = \mathbb{R}e_1, \quad W = \mathbb{R}(e_1 + e_2)$$

where $e_1 = (1, 0)$ and $e_2 = (0, 1)$. So, V_1 and W are one dimensional spaces. Then, with the choices

$$\varepsilon_0 = 1, \quad \varepsilon_1 = \frac{1}{2}, \quad w = \Big(\frac{\sqrt{3}+1}{4}, \frac{\sqrt{3}+1}{4}\Big),$$

it is easily seen that $\mathcal{K}_w$ is the single point $\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$ and has therefore null Chebyshev radius while $\mathcal{K}_w^0$ and $\mathcal{K}_w^1$ have positive Chebyshev radii.

In more general settings we do not have such a simple description of $\mathcal{K}_w$, however we now give some additional examples that show that even the a priori estimates of the previous section can be significantly better than the one space estimate as well as the estimate (3.5). We consider the two extremes in the compatibility between the favorable basis $\{\phi_1^*, \dots, \phi_n^*\}$ and the basis $\{\phi_1, \dots, \phi_n\}$ which describes the approximation properties of the sequence $\{V_0, \dots, V_n\}$.

Example 1: In this example we consider the case where the two bases coincide,

$$\phi_i^* = \phi_i, \quad i = 1, \dots, n.$$

Note that in this case the singular values $\{s_1, \dots, s_k\}$ for the pair $\{V_k, W\}$ coincide with the first k singular values for the pair $\{V_n, W\}$. Therefore

$$\mu(V_k, W) = s_k^{-1}, \quad k = 0, \dots, n,$$

where we have set $s_0 := 1$. We also have

$$\theta_k = \varepsilon_{k-1}, \quad k = 1, \dots, n.$$

We fix $\varepsilon_n := \varepsilon$ and $\varepsilon_{n-1} := \varepsilon_{n-2} := \varepsilon^{1/2}$ and the values $s_n := \varepsilon$ and $s_{n-1} := s_{n-2} := \varepsilon^{1/2}$ and all other $\varepsilon_k := 1$ and all other $s_k := 1$. We examine what happens when ε is very small. The estimate (1.9) would give the bound

$$\min_{0 \le k \le n} \mu(V_k, W)\varepsilon_k = \min_{0 \le k \le n} s_k^{-1}\varepsilon_k = 1,$$

as the bound for $\mathrm{rad}(\mathcal{K}_0)$ and $E(\mathcal{K})$. On the other hand, since,

$$s_n^2 \varepsilon_{n-1}^2 = \varepsilon^3 \ll \varepsilon^2 \quad \text{and} \quad s_{n-1}^2 \varepsilon_{n-2}^2 = \varepsilon^2,$$

the value of k in (3.7) is $n - 1$. It follows that the error E_n in the multi-space method (3.9) satisfies

$$E_n^2 \le \varepsilon_{n-2}^2 + \varepsilon_{n-1}^2 + \varepsilon_n^2 \le 3\varepsilon.$$

Hence, the error for the multi-space method can be arbitrarily small as compared to the error of the one-space method.

Example 2: We next consider the other extreme where the two bases are incoherent in the sense that each entry in the change of basis matrix satisfies

$$|\lambda_{i,j}| \leq C_0 n^{-1/2}, \quad 1 \leq i,j \leq n.$$

We want to show that E_n can be smaller than the estimate in (3.5) in this case as well. To illustrate how the estimates go, we assume that $n \geq 2$ and $|\lambda_{i,j}| = 1/\sqrt{n}$, for all $1 \leq i,j \leq n$. We will take

$$s_n \ll s = s_1 = s_2 = \ldots = s_{n-1},$$

with the values of s and s_n specified below. We define

$$\varepsilon_0 := 1/2 \quad \text{and} \quad \varepsilon_j = \frac{1}{2(n-1)}, \quad j = 1, \ldots, n-1,$$

so that $\sum_{j=0}^{n-1} \varepsilon_j = 1$. It follows from the definiton of θ_k given in (3.6) that

$$\theta_k = 1/\sqrt{n} := \theta, \quad k = 1, \ldots, n.$$

With these choices, the best one space estimate (1.9) is

$$\min\{\varepsilon_0, s^{-1}\varepsilon_{n-1}, s_n^{-1}\varepsilon_n\}. \tag{3.10}$$

Now, we take ε_n very small and $s_n = \varepsilon_n^2$. We then choose s so that

$$(s^2 + s_n^2)\theta^2 = \varepsilon_n^2. \tag{3.11}$$

This gives $k = n-1$ in (3.7) and so

$$E_n^2 = \varepsilon_n^2 + \theta_{n-1}^2 + \theta_n^2 \leq 3n^{-1}.$$

On the other hand, (3.11) says that $s^{-1} = \varepsilon_n^{-1}(n - \varepsilon_n^2)^{-1/2}$. Thus, from (3.10), the best one space estimate is

$$\min\{\varepsilon_0, s^{-1}\varepsilon_{n-1}, s_n^{-1}\varepsilon_n\} = \min\left\{\frac{1}{2}, \frac{1}{2(n-1)\sqrt{n-\varepsilon_n^2}}\varepsilon_n^{-1}, \varepsilon_n^{-1}\right\} = 1/2,$$

provided $\varepsilon_n \leq n^{-3/2}$. Hence, the multi-space estimate (3.9) is better than the one space estimate by at least the factor $n^{-1/2}$ in this case.

3.3 Numerical algorithms

In this section, we discuss some possible numerical algorithms, based on convex optimization, for the multi-space case. For any given data $w \in W$, such that $\mathcal{K}_w$ is not empty, these algorithms produce, in the limit, an element $A(w)$ which belongs to $\mathcal{K}_w$, so that they are near optimal in the sense of (3.1) and (3.2).

We recall that $\mathcal{K}_w$ is given by

$$\mathcal{K}_w = \mathcal{H}_w \cap \mathcal{K}^0 \cap \mathcal{K}^1 \cap \cdots \cap \mathcal{K}^n.$$

One first observation is that although the set $\mathcal{K}_w$ may be infinite dimensional, we may reduce the search for an element in $\mathcal{K}_w$ to the finite dimensional space

$$\mathcal{F} := V_n + W,$$

which has dimension $d = m + n - p$, where $p = \dim(V_n \cap W)$. Indeed, if $u \in \mathcal{K}_w$, then its projection $P_\mathcal{F} u$ onto $\mathcal{F}$ remains in $\mathcal{K}_w$, since $u - P_\mathcal{F} u \in W^\perp \cap V_n^\perp$ implies

$$P_W P_\mathcal{F} u = P_W u = w,$$

and

$$\operatorname{dist}(P_\mathcal{F} u, V_j) \leq \operatorname{dist}(u, V_j) \leq \varepsilon_j, \quad j = 0, \dots, n.$$

Therefore, without loss of generality, we may assume that

$$\mathcal{H} = \mathcal{F},$$

and that the sets $\mathcal{H}_w$ and $\mathcal{K}^j$ that define $\mathcal{K}_w$ are contained in this finite dimensional space.

The problem of finding a point in the intersection of convex sets is sometimes referred to as *convex feasibility* and has been widely studied in various contexts. We refer to [9, 10] for surveys on various possible algorithmic methods. We restrict our discussion to two of them which have very simple expressions in our particular case. Both are based on the orthogonal projection operators onto the spaces $\mathcal{H}_w$ and $\mathcal{K}^j$. Let us first observe that these projections are very simple to compute. For the projection onto $\mathcal{H}_w$, we use the orthonormal basis $\{\omega_1, \dots, \omega_m\}$ of W. For any $u \in \mathcal{F}$ we have

$$P_{\mathcal{H}_w} u = P_{W^\perp} u + w = u - \sum_{i=1}^m \langle u, \omega_i \rangle \omega_i + w. \tag{3.12}$$

For the projection onto $\mathcal{K}^j$, we extend the basis $\{\phi_1, \dots, \phi_n\}$ into an orthonormal basis $\{\phi_1, \dots, \phi_d\}$ of $\mathcal{F}$. We then have

$$P_{\mathcal{K}^j} u = \sum_{i=1}^j \langle u, \phi_i \rangle \phi_i + \alpha\Big(\sum_{i=j+1}^d \langle u, \phi_i \rangle \phi_i \Big), \quad \alpha := \min\Big\{1, \varepsilon_j \Big(\sum_{i=j+1}^d |\langle u, \phi_i \rangle|^2 \Big)^{-1/2} \Big\}.$$

We now describe two elementary and well-known algorithms.

Algorithm 1: sequential projections. This algorithm is a cyclical application of the above operators. Namely, starting say from $u^0 = w$, we define for $k \geq 0$ the iterates

$$u^{k+1} := P_{\mathcal{K}^n} P_{\mathcal{K}^{n-1}} \cdots P_{\mathcal{K}^1} P_{\mathcal{K}^0} P_{\mathcal{H}_w} u^k.$$

We know from general results on alternate projections onto convex sets [4] that this sequence converges towards a point $u^* \in \mathcal{K}_w$ when $\mathcal{K}_w$ is not empty. We make further use of the following observation: the nestedness property $V_0 \subset V_1 \subset \ldots \subset V_n$ implies that u^k belongs to $\mathcal{K} = \mathcal{K}^0 \cap \ldots \cap \mathcal{K}^n$.

Algorithm 2: parallel projections. This algorithm combines the projections onto the sets $\mathcal{K}$ according to

$$u^{k+1} := P_{\mathcal{H}_w} \Big(\sum_{j=0}^n \gamma_j P_{\mathcal{K}^j} \Big) u^k,$$

where the weights $0 < \gamma_j < 1$ are such that $\gamma_0 + \cdots + \gamma_n = 1$, for example $\gamma_j := \frac{1}{n+1}$. It may be viewed as a projected gradient iteration for the minimization over $\mathcal{H}_w$ of the differentiable function

$$F(u) := \sum_{j=0}^{n} \gamma_j F_j(u), \quad F_j(u) := \frac{1}{2}\,\mathrm{dist}(u, \mathcal{K}^j)^2.$$

Notice that the minimum of F is attained exactly at each point of $\mathcal{K}$. Since $\nabla F_j(u) = u - P_{\mathcal{K}^j} u$, we find that

$$u^{k+1} = P_{\mathcal{H}_w}(u^k - \nabla F(u^k)).$$

Classical results on constrained minimization methods [15] show that this algorithm converges toward a minimizer u^* of $F(u)$ over $\mathcal{H}_w$ which clearly belongs to $\mathcal{K}_w$ when $\mathcal{K}_w$ is not empty.

3.4 A posteriori estimate and convergence rates

Each of the above algorithms generates a sequence $(u^k)_{k\geq 1}$ of elements from $\mathcal{F}$ which are guaranteed to converge to a point in $\mathcal{K}_w$ provided that this set is nonempty. We would like to have a bound for $\mathrm{dist}(u^k, \mathcal{K}_w)$, since this would allow us to check the progress of the algorithm and also could be utilized as a stopping criterion when we have gained sufficient accuracy. Here we restrict our analysis to Algorithm 1.

We will use certain geometric properties of the set $\mathcal{K}$, expressed by the following lemma.

Lemma 3.3 *If $u_1, u_2 \in \mathcal{K}$ then the ball $B := B(u_0, r)$ centered at $u_0 := \frac{1}{2}(u_1 + u_2)$ of radius*

$$r := \frac{1}{8} \min_{j=0,\dots,n} \varepsilon_j^{-1} \|P_{V_j^\perp}(u_1) - P_{V_j^\perp}(u_2)\|^2 \tag{3.13}$$

is completely contained in $\mathcal{K}$.

Proof: For $u_1, u_2 \in \mathcal{K}^j$ the ball $B(u_0, r)$ is contained in $\mathcal{K}^j$ if and only if the ball in $V_j^\perp$ centered at $P_{V_j^\perp} u_0$ with the radius r is contained in $P_{V_j^\perp}(\mathcal{K}^j) = \{x \in V_j^\perp \ : \ \|x\| \leq \epsilon_j\} := \mathcal{B}_j$. Let $v_s^j := P_{V_j^\perp}(u_s)$ for $s = 0, 1, 2$ and let $\delta_j := \|v_1^j - v_2^j\|$. The parallelogram identity gives

$$\|v_0^j\|^2 = \frac{1}{2}\|v_1^j\|^2 + \frac{1}{2}\|v_2^j\|^2 - \frac{1}{4}\|v_1^j - v_2^j\|^2,$$

so that $\|v_0^j\|^2 \leq \varepsilon_j^2 - \frac{1}{4}\delta_j^2$. Thus for

$$r_j := \varepsilon_j - \sqrt{\varepsilon_j^2 - \frac{1}{4}\delta_j^2} = \varepsilon_j \left(1 - \sqrt{1 - \frac{\delta_j^2}{4\varepsilon_j^2}}\right),$$

the ball in $V_j^\perp$ centered at v_0^j with radius r_j is contained in $\mathcal{B}_j$. Thus, with

$$\rho := \min_{j=0,1,\dots,n} r_j,$$

we have $B(u_0, \rho) \subset \mathcal{K}$. Since $\delta_j \leq 2\varepsilon_j$ and $(1 - \sqrt{1-x}) \geq x/2$ for $0 \leq x \leq 1$ we get $r_j \geq \delta_j^2/(8\varepsilon_j)$ and therefore $\rho \geq r$ from which (3.13) follows. $\square$

We have noticed that the iterates u^k of Algorithm 1 all belong to $\mathcal{K}$ and we would like to estimate their distance from the convex set $\mathcal{K}_w$. Let $P_{\mathcal{K}_w}(x)$ denote the point from $\mathcal{K}_w$ closest to x. This is a well defined map. The following result gives an estimate for the distance of any $u \in \mathcal{K}$ from $\mathcal{K}_w$, in terms of its distance from the affine space $\mathcal{H}_w$. This latter quantity is easily computed using (3.12) which shows that

$$u - P_{\mathcal{H}_w} u = P_W u - w = \sum_{i=1}^{m} \langle u, \omega_i \rangle \omega_i - w.$$

Lemma 3.4 *Let $u \in \mathcal{K}$ be such that*

$$\alpha := \operatorname{dist}(u, \mathcal{H}_w) > 0.$$

Then

$$\|P_{\mathcal{H}_w} u - P_{\mathcal{K}_w} u\| \leq \rho = \rho(\alpha) := \max_j \mu_j (\alpha + 4\sqrt{\alpha \varepsilon_j}), \tag{3.14}$$

where $\mu_j = \mu(V_j, W)$. Since $u - P_{\mathcal{H}_w} u$ is orthogonal to $P_{\mathcal{H}_w} u - P_{\mathcal{K}_w} u$, we have

$$\operatorname{dist}(u, \mathcal{K}_w)^2 \leq \alpha^2 + \rho(\alpha)^2.$$

Proof: We set $u_2 = P_{\mathcal{K}_w} u$ and $\eta = u - u_2$ which we decompose as

$$\eta = (u - P_{\mathcal{H}_w} u) + (P_{\mathcal{H}_w} u - u_2) =: \eta_1 + \eta_2.$$

We wish to show that $\|\eta_2\| \leq \rho$, where ρ is defined in (3.14). To this end, observe that $\eta_1 \in W$ and $\eta_2 \in W^\perp$ so that this is an orthogonal decomposition. Moreover, using (1.5) and noting that $\|\eta_1\| = \alpha$, we have

$$\|P_{V_j^\perp} \eta\| \geq \|P_{V_j^\perp} \eta_2\| - \|P_{V_j^\perp} \eta_1\| \geq \beta(V_j, W) \|\eta_2\| - \alpha. \tag{3.15}$$

We infer from Lemma 3.3 that the ball B with center at $u_0 = \frac{1}{2}(u + u_2)$ and radius

$$r = \frac{1}{8} \min_{j=0,1,\dots,n} \varepsilon_j^{-1} \|P_{V_j^\perp} \eta\|^2$$

is contained in $\mathcal{K}$. Let us suppose now that $\|\eta_2\| > \rho$ and derive a contradiction. Then, we obtain from (3.15)

$$\|P_{V_j^\perp} \eta\| > \mu_j^{-1} \rho - \alpha \geq \mu_j^{-1} \mu_j (\alpha + 4\sqrt{\alpha \varepsilon_j}) - \alpha = 4\sqrt{\alpha \varepsilon_j},$$

and thus

$$r > \frac{1}{8} \min_{j=0,1,\dots,n} \varepsilon_j^{-1} 16 \alpha \varepsilon_j = 2\alpha.$$

On the other hand, note that $\|u_0 - P_{\mathcal{H}_w} u_0\| = \frac{1}{2} \|u - P_{\mathcal{H}_w} u\| = \alpha/2$. Therefore, $P_{\mathcal{H}_w} u_0 \in \mathcal{K}$ and hence in $\mathcal{K}_w$. Moreover,

$$\|u - P_{\mathcal{H}_w} u_0\|^2 = \alpha^2 + \frac{1}{4} \|u_2 - P_{\mathcal{H}_w} u\|^2,$$

and

$$\|u - u_2\|^2 = \alpha^2 + \|u_2 - P_{\mathcal{H}_w} u\|^2.$$

If $u_2 \neq P_{\mathcal{H}_w} u$, we have $\|u - P_{\mathcal{H}_w} u_0\| < \|u - u_2\|$ which is a contradiction since u_2 is the closest point to u in $\mathcal{K}_w$. If $u_2 - P_{\mathcal{H}_w} u = 0$ then $\eta_2 = 0$ contradicting $\|\eta_2\| > \rho$. This completes the proof. □

One immediate consequence of the above lemma is an a posteriori error estimate for the squared distance to $\mathcal{K}_w$

$$\delta_k := \operatorname{dist}(u^k, \mathcal{K}_w)^2,$$

in Algorithm 1. Indeed, we have observed that $u^k \in \mathcal{K}$ and therefore

$$\delta_k \leq \alpha_k^2 + \rho(\alpha_k)^2, \quad \alpha_k := \operatorname{dist}(u^k, \mathcal{H}_w).$$

This ensures the following accuracy with respect to the unknown $u \in \mathcal{K}_w$:

$$\|u - u^k\| \leq \sqrt{\alpha_k^2 + \rho(\alpha_k)^2} + 2 \operatorname{rad}(\mathcal{K}_w).$$

If we have an a priori estimate for the Chebyshev radius of $\mathcal{K}_w$, such as the bound E_n from Theorem 3.2, one possible stopping criterion is the validity of

$$\sqrt{\alpha_k^2 + \rho(\alpha_k)^2} \leq E_n.$$

This ensures that we have achieved accuracy $\|u - u^k\| \leq 3E_n$, however note that E_n can sometimes be a very pessimistic bound for $\operatorname{rad}(\mathcal{K}_w)$ so that significantly higher accuracy is reachable by more iterations.

We can also use Lemma 3.4 to establish a convergence estimate for δ_k in Algorithm 1. For this purpose, we introduce the intermediate iterates

$$u^{k+\frac{1}{2}} := P_{\mathcal{H}_w} u^k,$$

and the corresponding squared distance

$$\delta_{k+\frac{1}{2}} := \operatorname{dist}(u^{k+\frac{1}{2}}, \mathcal{K}_w)^2.$$

Since the distance to $\mathcal{K}_w$ is non-increasing in each projection steps, it follows that

$$\delta_{k+1} \leq \delta_{k+\frac{1}{2}} = \delta_k - \alpha_k^2.$$

On the other hand, it easily follows from Lemma 3.4 that

$$\delta_k - \alpha_k^2 \leq \rho(\alpha_k)^2 \leq A\alpha_k,$$

where A is a constant depending on ϵ_j's, μ_j's and $\|u\|$. It is easily seen that this implies the validity of the inequality

$$\alpha_k \geq \sqrt{\delta_k + A^2/4} - A/2 \geq \frac{\delta_k}{\sqrt{A^2 + 4\delta_k}} \geq \frac{\delta_k}{\sqrt{A^2 + 4\delta_0}} := c\delta_k,$$

and therefore

$$\delta_{k+1} \le \delta_k - c^2\delta_k^2.$$

From this, one finds by induction that

$$\delta_k \le Ck^{-1}, \quad k \ge 1,$$

for a suitably chosen constant $C := \max\{c^{-2}, \delta_1\}$ taking into account that for any $t \ge 1$

$$\frac{C}{t}\Big(1 - \frac{Cc^2}{t}\Big) \le C\Big(\frac{t-1}{t^2}\Big) \le \frac{C}{t+1}\,.$$

□

Remark 3.5 *The above convergence rate $\mathcal{O}(k^{-1/2})$ for the distance between u^k and $\mathcal{K}_w$ is quite pessimistic, however, one can easily exhibit examples in which it indeed occurs due to the fact that $\mathcal{H}_w$ intersects $\mathcal{K}$ at a single point of tangency. On the other hand, one can also easily find other examples for which convergence of Algorithm 1 is exponential. In particular, this occurs whenever $\mathcal{K}_w$ has an element lying in the interior of $\mathcal{K}$.*

References

[1] O. Bashir, O Ghattas, J. Hill, B. Van Bloemen Waanders, K. Willcox, *Hessian-based model reduction for large-scale data assimilation problems*, in Computational Science – ICCS 2007, Springer, Lecture Notes in Computer Science **4487**, 1010-1017, 2007.

[2] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Convergence Rates for Greedy Algorithms in Reduced Basis Methods*, SIAM Journal of Mathematical Analysis **43**, 1457-1472, 2011.

[3] B. Bojanov, *Optimal recovery of functions and integrals.* First European Congress of Mathematics, Vol. I (Paris, 1992), 371-390, Progr. Math., 119, Birkhuser, Basel, 1994.

[4] L.M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Computational Mathematics and Mathematical Physics **7**, 200-217, 1967.

[5] A. Buffa, Y. Maday, A.T. Patera, C. Prud'homme, and G. Turinici, *A Priori convergence of the greedy algorithm for the parameterized reduced basis*, Mathematical Modeling and Numerical Analysis **46**, 595-603, 2012.

[6] A. Cohen and R. DeVore, *Approximation of high dimensional parametric pdes*, to appear in Acta Numerica, 2015.

[7] A. Cohen, R. DeVore and C. Schwab, *Analytic Regularity and Polynomial Approximation of Parametric Stochastic Elliptic PDEs*, Analysis and Applications **9**, 11-47, 2011.

[8] A. Chkifa, A. Cohen, R. DeVore, and C. Schwab, *Sparse Adaptive Taylor Approximation Algorithms for Parametric and Stochastic Elliptic PDEs*, M2AN **47**, 253-280, 2013.

[9] P.L. Combettes, *The convex feasiblility problem in image recovery*, in: Advances in imaging an electron physics, **85** , 155-270, Academic Press, New York, 1996.

[10] P.L. Combettes and J.C. Pesquet, *Proximal splitting methods in signal processing*, in: Fixed-point algorithms for inverse problems in science and engineering, 185-212, Springer Verlag, New York, 2011.

[11] R. DeVore, G. Petrova, and P. Wojtaszczyk, *Greedy algorithms for reduced bases in Banach spaces*, Constructive Approximation, **37**, 455-466, 2013.

[12] R. Daley, *Atmospheric Data Analysis*, Cambridge University Press, 1991.

[13] M. von Golitschek, G. Lorentz, and Y. Makovos, *Constructive Approximation, vol II*, Springer Verlag, 1996.

[14] J. M. Lewis, S. Lakshmivarahan, S. Dhall, *Dynamic Data Assimilation : A Least Squares Approach*, Encyclopedia of Mathematics and its Applications 104, Cambridge University Press, 2006.

[15] E.S. Livitin and B.T. Polyak, *Constrained minimization methods*, USSR Comput. Math. Phys. **6**, 1-50, 1966.

[16] Y. Maday, A.T. Patera, J.D. Penn and M. Yano, *A parametrized-background data-weak approach to variational data assimilation: Formulation, analysis, and application to acoustics*, Int. J. Numer. Meth. Eng., submitted, 2014.

[17] C.A. Micchelli, T.J. Rivlin, *Lectures on optimal recovery.* Numerical analysis, Lancaster 1984 (Lancaster, 1984), 21-93, Lecture Notes in Math., **1129**, Springer, Berlin, 1985.

[18] C.A. Micchelli, T.J. Rivlin, and S. Winograd, *The optimal recovery of smooth functions*, Numerische Mathematik **26**, 191-200, 1976.

[19] P. Wojtaszczyk *On greedy algorithm approximating Kolmogorov widths in Banach spaces*, J. Math. Anal. Appl. **424**, 685-695, 2015.

Peter Binev
Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA
binev@math.sc.edu

Albert Cohen
Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie, 75005, Paris, France
cohen@ann.jussieu.fr

Wolfgang Dahmen
Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, D-52056 Aachen, Germany
dahmen@igpm.rwth-aachen.de

Ronald DeVore
Department of Mathematics, Texas A&M University, College Station, TX 77840, USA
rdevore@math.tamu.edu

Guergana Petrova
Department of Mathematics, Texas A&M University, College Station, TX 77840, USA
gpetrova @math.tamu.edu.

Przemyslaw Wojtaszczyk
Interdisciplinary Center for Mathematical and Computational Modelling University of Warsaw, 00-838 Warsaw, ul. Prosta 69, Poland
wojtaszczyk@icm.edu.pl